DOCUMENT RESUME

ED 144 978 TM 006 527

TITLE Symposium: Perspectives on Formative Evaluation of Children's Television Programs.
PUB DATE 6 Apr 77
NOTE 46p.; Papers presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Childrens Television; *Educational Television; Elementary Secondary Education; Field Studies; Formative Evaluation; Minority Group Children; National Programs; *Program Design; *Program Evaluation; Regional Programs

ABSTRACT

Evaluators of television programing and representatives of funding agencies discussed the impact of the perceptions of funding agencies on the evaluation of children's television. Participants also examined the interplay between the objectives of the television series and the evaluation, the relationship between production and evaluation, and the resulting problems. Participants and their topics included: Valerie Crane, Designing an Evaluation of a Regional Television Series; Francis Harvey, Designing Evaluation Programs to Meet the Changing Needs of a National Children's Television Series; Mel M. Moore, The Role of Evaluation in the Production of Children's Television Programs; and Saul Rockman, Evaluation of an Instructional Television Series. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144978

American Educational Research Association
1977 Annual Meeting
New York, New York

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Symposium: Perspectives on Formative Evaluation of Children's Television Programs

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Valerie Crane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

April 6, 1977

Organizer: Valerie Crane, Ph.D., Heuristics, Inc.

Chairperson: Robert Consalvo, Ph.D., Heuristics, Inc.

Participants: Valerie Crane, Ph.D., Heuristics, Inc.
Mel Moore, WGBH-TV
Frank Harvey, Education Development Center
Saul Rockman, Agency for Instructional Television

Discussant: Joanne Linowes, Massachusetts Educational Television

TM006 527

PERSPECTIVES ON FORMATIVE EVALUATION OF

CHILDREN'S TELEVISION PROGRAMS

American Educational Research Association

Chairperson: Consalvo, Robert, Ph. D.
Heuristics, Inc.
141 Linden Street
Wellesley, Massachusetts 02181

"Designing an Evaluation of a Regional Television Series"
Crane, Valerie, Ph. D.
Heuristics, Inc.
141 Linden Street
Wellesley, Massachusetts 02181

"Designing Evaluation Programs to Meet the Changing Needs of a National Children's Television Series"
Harvey, Francis
Education Development Center
55 Chapel Street
Newton, Massachusetts 02160

"The Role of Evaluation in the Production of Children's Television Programs"
Moore, Mel M.
WGBH-TV
125 Western Avenue
Boston, Massachusetts 02134

"Evaluation of an Instructional Television Series"
Rockman, Saul
Agency for Instructional Television
Box A
Bloomington, Indiana 47401

Discussant: Linowes, Joanne
Massachusetts Educational Television
54 Rindge Avenue Extension
Cambridge, Massachusetts 02140

PERSPECTIVES ON FORMATIVE EVALUATION OF CHILDREN'S TELEVISION PROGRAMS

Organizer: VALERIE CRANE, Ph. D.
Heuristics, Inc.

Chairperson: ROBERT CONSALVO, Ph. D.
Heuristics, Inc.

Introduction

A variety of professionals, representing different areas of expertise, become involved in the development of children's television programs. Educators become involved as content directors and evaluators; producers obviously specialize in the technology of television; and those representing the funding source combine expertise in these areas with administrative skills. Due to the relatively recent development of evaluating children's programs, there have been few opportunities for professionals to discuss their perspectives on the formative evaluation of children's television programs.

This symposium brings the evaluators and those representing the funding source together in order to discuss the following topics.

The Scope of the Evaluation Effort

- How do the perceptions of the funding source affect the evaluation effort?
- How do the objectives of the television series affect the evaluation effort?

The Relationship between Production and Evaluation

- What kind of relationship exists between production and evaluation?
- What are some of the problems found in the relationship between production and evaluation?

The formative evaluation of children's television programs plays an ongoing role in determining the effectiveness of a television series. The production of a series involves a substantial investment of funds, and accountability for this investment is of critical importance in order to assure program success.

The educational potential of television has been well-documented in the educational, psychological, and now in the medical literature. The Office of Education recognizes the educational role of television both in and out of school and supports a number of television programs under the Emergency School Aid Act. The primary focus of this programming is to provide quality home-viewing programs developed by minority personnel for minority children. Mel Moore from WGBH-TV, Frank Harvey from Education Development Center, and Valerie Crane from Heuristics, Inc. have presented papers on the evaluations of two national and one regional ESAA television series.

The educational impact of a series is of particular concern to producers of programs designed specifically for classroom use. Evaluation of instructional television programs is usually confined to the formative stages of production; rarely, if ever, is evaluation of educational impact conducted by those agencies

which purchase the programs. Saul Rockman from the Agency for Instructional Television will discuss the importance of formative evaluation in the development of an instructional television series.

The participants in the symposium examine the relationship between production and evaluation. The success of this relationship directly affects the degree to which the evaluation findings can be incorporated into program content and development. On yet another level, decisions are made by the funding source or by broadcasters based upon the evaluation findings. Often, a pilot program is evaluated before an entire series is produced. This type of formative evaluation forms the basis for the decision to continue or discontinue production. Discussant Joanne Linowes from Massachusetts Educational Television will present her views of the relationship between the evaluator and funding source as well as the role of evaluation in the decision-making process.

Designing the Evaluation of a Regional Television Series

Valerie Crane
Heuristics, Inc.

The topic for my presentation is a regional television series which is in the early stages of development, and where planning of production, content, and evaluation activities are currently taking place. This regional television series is about Native Americans in Southern New England. The program will be developed for home viewing as well as for use in the schools and will be produced by Massachusetts Educational Television in conjunction with WGBY of Springfield, Massachusetts. The main objectives of the series are to increase knowledge and appreciation of Native American culture and to increase positive attitudes toward Native Americans among both Native American and non-Native American students in grades four through six.

Since we are currently designing the evaluation of this series, I would like to elaborate on some of the problems encountered from the outset as the evaluation takes shape. The first set of issues affecting the evaluation effort relate to the requirements specified by the funding source, and second set of issues represents problems of methodology which are faced by all evaluators involved in the evaluation of television programs.

First, I will discuss some of the requirements imposed on the evaluation effort by the funding source. According to the RFP issued by ESAA,

> Formative evaluation will be limited to an evaluation of the pilot program(s). Primary concern of the evaluation design should be to determine the appeal and attention hold of the pilot(s). Data on cognitive gains, to the degree that meaningful data can be obtained from a pilot program, should also be included.
>
> High points will be awarded for proposed testing techniques which can achieve the above goals and for the adequacy of the sample to be employed in the testing, both in terms of its size and of its geographical and ethnic representativeness.

As both budget and scheduling are considered, it becomes apparent that the evaluation will be performed when a pilot program is completed. Upon completion of the evaluation report, both the pilot program and evaluation report are sent to Washington and during a review period all production halts while the continuation of production is considered by ESAA staff. This evaluation report is not prepared primarily for the production and content teams but rather for policy and decision makers at the administrative level. In effect, the evaluator has conducted a summative, not formative, evaluation. It can be observed that by limiting the evaluation to an already completed pilot, that the evaluation cannot have ongoing input into program development. It has been my experience that such a division of services severely reduces the effectiveness of production and hampers the necessary positive relationship between production and evaluation. The evaluators are put in the position not of helping production of the pilot but of "judging" it.

On the Native American Project we have decided that an ongoing relationship with production is necessary in order to assist both producers and content staff in developing and adhering to objectives which can be translated into film treatment and subsequently measured.

An ongoing relationship is also important with the Advisory Board--a group of professional and community members who oversee all aspects of the project. On the Native American Project, the Advisory Board has offered help in the areas of instrument development (with special attention given to stereotyping and language problems) and locating the subjects for the study (a considerable task since the project is being evaluated during the summer.)

In addition to those already mentioned, other problems result from the evaluation of a single pilot program. There is no absolute standard of a successful television program. The level of sophistication in the evaluation of television is not sufficient to warrant our talking about norms or predetermined levels of performance. A more effective approach in determining the effectiveness of a program would be comparative evaluation. Determining which film treatment or which content approach produces greater attention, appeal, and comprehension would produce more meaningful data than determining what percent of the students attended, liked, or understood the material. In fact, to claim that a program commands a high rate of attention and has a high degree of appeal are deceptive statements. The evaluator goes into a classroom and asks the students to watch a television program. We faithfully record eye contact and then ask them if they liked what they saw. The hidden comparison is between the teacher and the television program, and not surprisingly, the program always wins hands down. For this phase of evaluation, it would appear that comparison of two programs on the same topic might produce more valid results.

As the evaluation of the pilot program is refined, a variety of methodological concerns arise. I would like to discuss some problems encountered related to the selection of the sample, the development of valid measures appropriate for the intended viewers, and the setting for the field test.

Obviously, the sample must be representative of the target audience if results are to be generalized beyond the evaluation effort. This regional series is intended for students representing an age range spanning three to four years. Special consideration must be given to age differences not only in the evaluation but in the production of program content.

In addition to age differences, we have to secure a representative sample from three different states. Conducting an evaluation in a number of states becomes a costly procedure and after the data has been gathered, there is the perplexing problem of whether to examine locale as a variable and how to interpret differences which can be attributed to various regions.

We also have a program designed for a multi-ethnic target audience. The required sample consists of 350 students, (175 Native American, 115 white, and 60 other minority). The evaluator is well aware that research has indicated that minority group children are affected not only by the race of the interviewer, examiner, or tester (Sattler, 1970) but by the language and culture bias of the test itself (Linden, Linden, & Bodine, 1974). The sampling and instrumentation are seen then to be interactive aspects of the evaluation design.

A second area of concern in evaluation is the development of valid measures, appropriate for the intended viewers. Here the relevant problems include the limited reading ability of students, language barriers among bilingual groups, and scheduling constraints that affect the kind of measures given to large groups.

In the evaluation of children's television, the measurement problem is complicated by the requirement of developing different instruments for each program. Tight production schedules do not allow for ample time to test, analyze results, retest, and further revise measures. Often the evaluator decides upon a test format on the basis of previous experience, a review of the few evaluation studies available, or on the basis of the budget. Common evaluative procedures include paper and pencil formats such as multiple-choice and true-false, the interview, rating scales, and the semantic differential. Formats are often only informally tested, and no validity or reliability estimates are reported. Seldom is there adequate data on the relative advantages of the different formats and procedures.

A third concern in the evaluation effort is the setting in which the viewing takes place. While the most manageable setting for evaluation is in the schools or recreational centers, it must be recognized that ESAA programs are primarily intended for home-viewing. Lyle and Hoffman (1972) have indicated that home-viewing patterns are quite different from school-viewing patterns. Attention patterns of viewing sometimes rate above 90% in a school setting, but the rates could be quite different at home. The relative effectiveness of the program in

the two settings must therefore be considered before any definitive statements about program effectiveness can be made.

At present, data obtained on one instrument in a certain setting may cause a program to appear more favorable than data from another, with resulting differences in decisions about program effectiveness. Such variability in decision criteria can effectively destroy the credibility of the evaluation effort. Evaluators need to know how data obtained with different formats, in a variety of settings, and among different ethnic groups can lead to different decisions, before they can establish realistic decision criteria.

It would appear that the problems in methodology outlined above should be addressed in a series of studies designed specifically to provide the evaluator with the necessary information to make systematic decisions about the evaluation design. We are talking about a very substantial investment of tax dollars in the development of these television programs. Therefore, a more sophisticated and systematic approach to evaluation is essential if we, as evaluators, are to increase our credibility with the funding source and production team with whom we should be working. Finally, formative evaluation must become an integral part of program development in order, first, to ensure more effective programming and, second, to achieve credibility with the toughest critics of all--the audience.

REFERENCES

Linden, K., Linden, J., & Bodine, R. Test bias: Fuss n' facts measurement and evaluation in guidance, 1974, 7, 163-168.

Lyle, J., & Hoffman, H. R. Children's use of television and other media. Vol. IV Television in day-to-day life: Patterns of use. In Television and social behavior. Washington, D. C.: Surgeon General's Report, 1972.

Sattler, J. M. Racial 'experimenter effects' in experimentation testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 3, 137-160.

Designing Evaluation Programs to Meet the Changing Needs of a National Children's Television Series

Francis A. Harvey
Education Development Center, Inc.

"Infinity Factory" is a new television series about people using mathematics to solve real-life problems. The series is produced by Education Development Center in Newton, Massachusetts, with funds from the U.S. Office of Education under the Emergency School Aid Act (ESAA). The programs are designed for both home and school viewing and are aimed at children 8-12 years old, with special emphasis on meeting the needs of Black and Latino children. The first half-hour "Infinity Factory" programs are now being aired on 175 PBS stations across the country and are viewed in about two million homes per week. Production of a second series of 30 half-hour programs, also funded by the U.S. Office of Education, began in February, 1977.

The evaluation programs for the first and second years of the "Infinity Factory" series are very different, reflecting the changing needs of the project at various stages of its development. This paper will discuss how the "Infinity Factory" evaluation program evolved over the past two years, and also how the past two years' experience has shaped what is now being proposed for evaluation during the third year of production.

"Infinity Factory I" was a new show which was designed to appeal to a national audience and at the same time promote math learning. The series

directed its appeal to a varied geographical and ethnic audience through a format which featured two dramatic segments aimed at particular segments of the target audience: "Scoops' Place", about a Black family running a neighborhood store in New York; and "City Flats", about a Latino family operating a bakery in East Los Angeles. Completing the magazine format program were "Brownstone" segments, featuring a multi-ethnic cast of young people in short skits that take place in and around an urban brownstone apartment house built in a television studio; animation segments; "Math in the Street" interviews; and historical segments about the contributions of notable minority persons.

The major questions which the project wanted the evaluation effort to attempt to answer concerned: 1) the overall appeal of the programs; 2) the appeal of particular major segments of the program; 3) whether children understood the dramatic story lines; 4) the amount of math learning which took place; 5) whether children's attitudes toward math, and a range of other attitudes related to the social and cultural goals of the series changed as a result of watching the programs; and 6) teachers' reactions to the programs. To answer these questions, an evaluation study[1] was designed which involved over 1,000 children and over 40 teachers in Boston, Massachusetts; Lawrence, Massachusetts; Los Angeles, California; and New York, New York. Students were pretested on knowledge of math content and social and math attitudes; the same instruments were administered as a posttest after the children had watched one "Infinity Factory" program per week for eight weeks. Children's attention was monitored during

program viewing, and children were tested immediately after each program for appeal, comprehension of dramatic story line, understanding of math content, and social and math attitudes. Teachers' attitudes toward math and television were also pre- and post-tested, and their opinions of each program measured each week.

No control group was used in the evaluation study because the evaluators did not have access to sufficient background data on the children in the sample (math test scores, records of which areas of the math curriculum the children had already been taught, etc.) to provide reliable matching of control and experimental groups, and because many of the research questions the evaluation attempted to answer (attention, appeal, comprehension of story line) did not require a control group.

Results of the evaluation study supported the conclusions that the "Infinity Factory I" programs were able to capture and hold children's attention; that the programs had high appeal for Black, Latino, and other children; that "Scoops' Place" and "City Flats" both appealed to both Black and Latino children and that teachers considered the programs effective for motivating children toward learning math and introducing math topics. The most significant finding of the study was that all student groups (younger and older children; boys and girls; and Black, Latino, and other children) showed significant gains from pretest to posttest in knowledge of math content. The gain scores for all groups were on the order of one standard deviation, and were considered large enough to be both educationally and statistically significant.

This study answered the major questions which the project needed answered at that stage in its development, and provided a considerable data base on which to base future production decisions. However, the amount of time needed to carry out such a study placed some limitations on its value for influencing on-going production. The processes of developing and refining instruments, administering the instruments, interpreting data collected, and reporting the results were going on at the same time as the development and production of the rest of the programs in the series. In fact, the final week of posttesting coincided with the final week of series production. Although preliminary and informal results were fed back to production and content staff as the show-by-show data were collected, the overall results could not be available in time to influence production of the "Infinity Factory I" series.

A major recommendation of the report of the eight-show evaluation was that evaluation efforts for the second year's production involve smaller groups of children, be more informal, and provide more immediate feedback. This recommendation is now being implemented. Production of the first "Infinity Factory II" programs is scheduled for February. "Infinity Factory II" will have the same basic format as the first 52 programs, with some important differences. The new series will feature four different families who are neighbors in an urban multicultural neighborhood: The Robinsons, a Black family; the Ramoses, a Puerto Rican family; the Garzas, a Chicano family; and the Mannings, a white family. The children of these families are close friends. All videotaping will

be done on location rather than in a television studio. These changes make possible a much greater amount of intercultural sharing not only among children from different ethnic groups but also among their families. The new programs will also include animation segments, film documentaries about how people use math in their professions, and songs sung by the children in the cast.

The process of gathering feedback about the new programs has already begun. Present plans call for gathering feedback in two different ways. First, a small, relatively controlled study will be carried out with groups of children in several different cities. This study will examine children's attention to the program while they view it, and then assess the appeal of the program overall and of major segments, comprehension of dramatic story lines, and learning of math content through a structured interview.

Second, groups of children and teachers who have been using the "Infinity Factory I" programs during the year will be asked to view the "Infinity Factory II" program and react to it via an open-ended response form. This will provide a wider range of (admittedly informal) responses, and at the same time provide some measure of how people familiar with the "Infinity Factory I" series react to the second year's programs.

Although the production schedule for the "Infinity Factory II" series is fairly tight, the evaluators will be able to begin developing instruments and selecting evaluation sites before final editing of the programs, so that the

evaluation can begin as soon as the final version of the first programs are available. The smaller numbers of subjects should make it easier to provide reliable and valid feedback soon enough to provide direction to ongoing program production.

Gathering feedback from children about the "Infinity Factory II" programs has been the responsibility of the math content team. However, while production is underway the demands on the math content people's time are very great. Among other things, they are responsible for explaining the math content to writers and production staff, discussing script ideas with writers, reviewing drafts of scripts, being present during actual videotaping or filming to ensure that the math ideas are presented correctly and clearly, consulting in post-production editing decisions, and developing program guides. As a result, it is often difficult to make time available for evaluation activities. Because the project recognizes that continuous interaction with and feedback from target audience children is essential to the success of the project, we are proposing that for "Infinity Factory III" a single staff member be given full-time responsibility for gathering informal data from children on an everyday basis, throughout all phases of planning, scripting, and production. Working closely with the content and production groups, the evaluator can provide data turnaround in a matter of hours on, for example, baseline mathematical skills, reactions to story lines and to individual segments, character preferences and identifications, clarity of mathematical presentations, and so on--whatever information the content and production groups need at the time.

This ongoing evaluation will not involve large numbers of children, or statistical analysis, or travel between cities; instead, it will provide fast, rough answers to very specific questions. For example, content staff often ask one another, "Which way of presenting this piece of mathematics will be clearest?" Production staff sometimes needs to know, "Will the kids understand the point of this story right away?" "Will anybody see this character as a role model?" With carefully chosen, small samples of children, a full-time evaluator can provide working answers to questions like these. While not absolutely reliable, the results should be an improvement on intuition.

The major goal of the "Infinity Factory" project is to make television programs that children will watch, enjoy and learn from. The addition of a formative evaluator/math content specialist to project staff will be of considerable help in meeting this goal.

NOTE

1. Harvey, Francis A., Barbara Quiroga, Valerie Crane, and Charles Bottoms. "Evaluation of Eight 'Infinity Factory' Programs." Newton, Mass.: Education Development Center, 1976. (ED 129 329 through ED 129 332).

The Role of Evaluation in the Production of Children's Television Programs

Melvin M. Moore
WGBH-TV

Within the past few years we have witnessed an unprecedented national effort to promote the social, cultural and intellectual development of children through the use of educational television programming. This effort to manipulate the impact of television on children has many sources but undoubtably the most significant is the recognition that television is a powerful and ubiquitous agent of socialization.

Concurrent with the tremendous concern with the effect of television content on children there has been some emphasis on evaluative research in the planning and production of children's television programs. However, while concern with impact or effect has led to a proliferation of children's educational programs, there has been less than comparable concern with the assessment of the impact of these programs; there appears to have been and continues to be little acceptance of the role of evaluation as an integral part of the production process in the creation of children's television programs.

This generalization is based primarily on the funding requirements of a federal agency that is a major funding source of many public television children's programs. In its request for proposals since 1974, applicants have been instructed to include formative evaluation as a part of their proposed series but are also advised to limit evaluation to the pilot program. Further, although

the programs to be funded are "educational" in nature--aimed at such things as "attitude change," "self-image," "cognitive growth," and "cultural awareness"--applicants have also been advised that the "primary concern of the evaluation design should be to determine the appeal and attention hold of the pilot". This concern ignores the issues of intended and unintended message transmission and content comprehensibility and effect. And, unfortunately, it addresses the objectives of commercial, more so than that of supposedly educational television productions.

This attitude toward evaluation deemphasizes the value and the role of evaluation in the production of children's television. Because this attitude is communicated either directly or indirectly to production personnel, it is often destructive of efforts to make high quality, highly entertaining programs, educationally significant as well.

Today, in spite of much public and professional attention being directed toward the content of children's television, few completely non-violent, non-racist, non-sexist or completely pro-social programs have been produced. There is therefore little justification for the under-utilization of evaluation in the development of positive program content.

However, justified or not, today many educational programs with high audience appeal are failing to achieve fully their educational objectives, as producers are often without sufficient feedback from their audience, and are therefore unable to objectively appraise their work.

As a case in point, I will review the evaluation activities of a national children's series during its first two years. That program is REBOP, produced by WGBH-TV in Boston. REBOP is a multi-cultural series under Title VII of the Emergency School Aid Act. Produced as a 26 half-hour semi-documentary series, each program consists of three nine-minute segments, each segment presenting a child of different ethnic background. The 78 children portrayed in the series were selected from minority cultures of the United States; namely Black, Latin, Asian and Native American.

As an educational series, REBOP was designed to promote, among 9-13 year olds, positive attitudes toward ethnic minority group cultures in American society.

Specifically, the series was proposed to:

a) deepen children's understanding of the particular contributions of different minority cultures to American life.

b) reinforce in minority-group children a positive sense of their own cultural heritage.

c) encourage in children a positive regard for others belonging to ethnic groups different from their own.

d) offer to minority-group children in particular, the references for positive self-image.

e) promote positive approaches to intergroup relations.

During the early phase of the project primary concern was focused on the production and evaluation of a semi-documentary program that would appeal

to nine through thirteen year olds. Evaluation during this period focused on the pilot's appeal--its ability to attract and sustain audience attention. Of specific interest was the relationship between appeal and certain viewer characteristics such as ethnicity, age, sex, and social status.

The sample consisted of 1,100 fourth through eighth grade boys and girls representing the age range of the intended viewing audience. Racially, the sample included Black and white American, Chinese American, Native American and Latin American children. Geographically, these children were recruited from school systems in Boston; Chicago; Austin, Texas; Akron, New York; Des Moines, Iowa; and San Francisco.

Two types of instruments were used during pilot evaluation: 1) observational and 2) questionnaire. They were both administered in schools during regular school hours.

The instrument used for collecting observational data was a modification of a small group observation instrument frequently used in research relating to children and television. This instrument was designed and scored to yield data concerning degrees or levels of attentiveness to video material. The observational instrument consisted of three identical forms, a separate one to be used for coding behavior during each of the three film segments that constituted the pilot. The observation sample consisted of 488 students, including students from every city and grade level and of every ethnic and age group included in the total evaluation sample.

The evaluation questionnaire was a 27 item structured instrument that heavily relied on recall, character comparisons and personal preferences to yield the data desired. This instrument was individually administered.

As an indirect measure of interest, students were asked to recall the name of the program, the names of children that appeared in the pilot and what they remembered about each. Students were also asked to indicate how much they liked or disliked the pilot, the individual segments of the pilot and music contained in each. Finally, due to our interest in any relationship between general viewing preference and pilot appeal, students were also asked to compare the pilot with their favorite television programs.

Formative evaluation after the pilot involved evaluations of individual film segments. The concern during this period was also with appeal but primarily with comprehension of film content. The assessment of both was particularly important given the nature of the series but also because of the difficulty in producing material that appeals to and is understood by both nine and thirteen year olds. It was therefore important to get some indication from our audience as to their understanding of story lines, the specific educational messages contained in each segment, and whether or not ethnicity was perceived as a salient or significant aspect of each segment. Additionally, there was an interest in discovering if produced material conveyed unintended messages.

In general, evaluation on REBOP extended beyond the actual task of testing a film and reporting results to the production staff. It involved several elements, all designed to promote a positive attitude toward evaluation and its acceptance as an important aspect of television production.

Before testing began and during its execution, the process of evaluation included educating the production staff concerning the concept, purpose and methodology of formative evaluation in the design of children's educational programs. During this period, research studies on children's television were summarized and distributed to the production staff in order to increase their familiarity with children's television research and to convey certain information and concepts that might be helpful in the development and production of future programming.

Prior to the evaluation of film, producers were encouraged to compose questions concerning their films. To the extent possible their questions were incorporated into the evaluation questionnaire. Producers were also asked to provide, for each film produced by their respective teams, story lines, the specific messages that each film was intended to convey, and the desired effects, if any, that the film was designed to have on viewers.

The research methods and procedures employed during the pilot evaluation indicated that the pilot had considerable appeal among all children sampled.

Results from the evaluations that followed were also positive with respect to appeal but more importantly, it was found that children had difficulty following story lines and understanding specific messages contained within story content.

As film production increased, production schedules hampered the dissemination of evaluation results. Consequently, the success of a film increasingly became defined by production's idea of what children liked and less by what they actually liked or understood. The production of each film technically and aesthetically, seemed to become more important than its impact or the extent to which it addressed program objectives.

Evaluation and content development activities ended with less than 30 percent of the project's films evaluated and very few evaluation findings or recommendations considered in the development and production of program film segments. To some extent this was due to a production schedule that was not conducive to meaningful evaluation activity, lack of understanding of the role of evaluation by production staff and little assistance by project management in building a working relationship between evaluation and production personnel.

It became clear at the end of two years that a pilot evaluation was conducted because it fulfilled a grant requirement. Afterwards, a program of evaluation seemingly served only to support REBOP's credibility as an educational television program; evaluation results were used either to obtain additional funding or in special reports to the program's Advisory Committee.

I have not intended this presentation to be an indictment of REBOP. The problems encountered in the administration of its program of evaluation were largely the result of the contractual requirements designed by the Department of Health, Education and Welfare and the attitude that it unwittingly encouraged toward evaluation. These problems are probably duplicated in many, if not a majority of projects in which the significance of evaluation is minimized.

Evaluation on all federally funded projects will continue to be plagued by problems as long as evaluation is treated as a necessary evil.

If Federal, State and private funding sources are to continue providing funds for the production of educational television programs they might do well to seriously question their commitment to truly educational television programming, require that grant applicants conduct formative evaluation throughout program production, require production schedules that allow adequate time for on-going evaluation and feedback, require quarterly reporting of evaluation results, and allow for an evaluation staff and administrative budget.

Finally, optimal effectiveness in the production of children's television programs can probably be best achieved through recognition that evaluation has a role not only in the assessment of program impact but also in the development of film content and through acceptance of the evaluator as an essential member of the production staff.

Evaluation of an Instructional Television Series

Saul Rockman
Agency for Instructional Television

The Agency for Instructional Television is an independent, nonprofit organization that works with education and television agencies in the United States and Canada. It has two major missions. One, it acquires and distributes a wide variety of television programs and related materials for use in schools. Two, AIT develops significant instructional television programming in cooperation with a consortium of local, state, and provincial agencies. These consortium agencies pool their financial and intellectual resources to create television series that meet their common, significant education needs. AIT has been developing consortia since 1969, and has involved from 13 to 45 agencies in each one.

The content of these television series has ranged from early childhood education to economics for the fifth and sixth grades to basic skills to emotional health. The topics and grade levels are selected in conjunction with the consortium agencies.

The need is identified within education and verified in the educational marketplace. The agencies that share a common need must back up their interest with financial commitment, the market test of the worth of an idea--democracy at work. Although evaluation plays a role in each phase of a project's

development and production, this session deals primarily with formative evaluation and I will limit myself to a discussion of one project to illustrate what we have been doing.

At AIT the evaluation and research policy stresses formative evaluation. Each project has a budget for the revision of materials in the formative phase. Sometimes a Catch-22 emerges; that is, a project can have 15% of its money set aside for evaluation and revision, which means that the more you evaluate and find things that need revision, the less money you have to continue revision and evaluation.

In one of AIT's current projects, the Essential Learning Skills Project, a full 20% of the projection budget is set aside for revision. Since the Agency for Instructional Television has not been accepting direct federal support (and therefore we have no federal restrictions), I feel very differently from the earlier speakers. I feel a great freedom and after listening to the previous speakers, I'm sure many of you can understand why that is.

I don't know if many of you are familiar with contemporary school television and that is why I have described AIT and a little bit of what it does. Let me illustrate our process with "Self Incorporated," a 15-program television/film series designed to stimulate classroom discussion of the critical issues facing early adolescents. The series is aimed at helping 11- to 13-year-olds cope with the physical, social, and emotional changes they are experiencing; i.e.,

sibling rivalry, prejudice, cliques, sexual identifications, menstruation, and so on. It doesn't shirk from these issues--the ones that made us all so uncomfortable many years ago. In "Self Incorporated," dramatic and realistic vignettes illustrate the problems that may occur during this age period. Solutions are not offered--the programs are open-ended and the design calls for classroom discussion to follow the viewing of the programs.

Formative evaluation activities, or perhaps as Jim Sanders would call them, preformative evaluation activities, for this project began much earlier than when the first foot of film was shot. AIT is a management agency and has no production facilities of its own; therefore, it contracts writing and film making with agencies around the United States and Canada. Since writing and production was going on some distance from our Bloomington, Indiana headquarters, we thought it was counterproductive to have a centralized formative evaluation activity going on--especially with the way the mails are.

In conjunction with each of the production agencies and their writers we assigned local evaluation consultants who were paid through the evaluation budget, not the production budget. They served as another pair of hands, feet, and eyes for the writers and producers at each site. These local consultants, with access to classrooms in their own areas, worked with students in many ways--to test out story concepts, ideas, scenarios, anything they could do to improve the appeal and salience of the script writers' work. They tried to improve dialogue,

and to get more of the child's reality into the work of the adult writer. They were always available by telephone and occasionally one-day turnaround for formative evaluation data from classroom to writer was therefore possible.

Let me illustrate how some of these local evaluation consultants work, because the stories are good. One writer was given the job of creating a script on the topic of failure--that kids in early adolescence sometimes experience failure in their life and sometimes it has very drastic effects on them, a cycle of depression and anger and full circle to adjustment. This writer had an idea but he really didn't know how failure was perceived by children in the target audience. So the local evaluation consultant went to her classroom and asked her students to write some essays on their experiences with failure. After two days, the kids started turning in their essays but one young woman demanded an extra day to finish hers because she thought it was so important. Well, when she turned it in a day late, the writer read it and loved it immediately--it fit his idea of how a program should be about failure. The young woman tried out for the cheerleading team and didn't make it while some of her friends did. The cycle of failure/rejection worked its way out through the summer and she was back, happy, in school again that Fall. We paid the young girl an honorarium for her script idea and went back to the class time and time again to work out some of the details and to check on dialogue, events, and so on. We decided to produce the program in that school building and the young woman tried out for her own part . . . and she didn't get it . . . but she did make cheerleader that year, so it ended well.

As you can see, these preformative evaluation activities can lead to many minor and major changes during the scripting and early production phases.

In addition, we are involved with our sponsors, the consortium agencies, in the field testing of the materials as they become available. We do not have access to classrooms, and in fact we didn't have access to televisions until recently (we're in the management business, as I've said). We have to rely on the consortium agencies to help us find the classrooms and teachers we need to gather information on our projects. The consortium agencies work under AIT's supervision but they are responsible for the selection of the sample (with our help and training) and for the conducting of evaluation activities and reporting the data back to us. These evaluation activities involve more than the traditional field test along the lines of summative evaluation. It is still truly formative evaluation for us. Any program in the field testing stage can, and will, be revised. We have thrown out entire programs, $20,000 or so down the drain, because they just did not work. We have revised programs after they have been shot, just because a line or a scene or a character did not work.

Let me quickly tell you some of the things we do to find out whether or not they work. The first thing I should mention is visual attention and appeal. Many people pay too much attention to attention data. Alone, such data is relatively meaningless. It is only useful as a diagnostic tool for the assessment of a particular program and in and of itself is not good data on which to base

program decisions. We believe that attention is a necessary but not sufficient condition for learning from the program. Just because someone isn't watching doesn't mean they aren't learning. Attention data should be used diagnostically and followed up to find out why a particular segment is not holding attention or is not working.

A second evaluation concern is students' comprehension of the program material and their attitudes about characters and concepts. Beyond the level of attention and appeal, comprehension data is easiest to gather and is perhaps the most important measure for a program, because if you don't understand what is going on, there is very little chance that you are going to get anything out of it. In a study by Friedlander and Wetstone a couple of years ago, they report that less than one-half of the material presented in the program was understood by the audience. That may say something about the way communicators are communicating.

In addition to comprehension, attention, and appeal data, we are very much concerned with the material's effect in the classroom--what happens when the program's over, when the teacher turns off the set? "Self Incorporated" was supposed to lead to discussions about the critical issues of early adolescence. We created observation instruments to deal with the climate of and concerns expressed in the classroom following the program. Post-viewing discussions are often necessary to clarify issues, to extend the instruction in some way. A close analysis of classroom interaction data following a program can help

lead to improved in-service activities for the series, can lead to more extensive and adequate teacher guide materials.

For "Self Incorporated" we extended our observation system used for "Inside/Out" and dealt with whether or not the discussion focused on the events in the program or the perceptions of the students. All of this material is in our final evaluation report, available through ERIC, so I won't spend more time on the details.

We also gather teacher attitude information about the use of the series which helps us plan for the utilization of it.

In case you get the idea that we are using and abusing the consortium agencies as a convenient sample pool, don't be mistaken. They gain a great deal from participation in the evaluation. Involvement in many ways is its own reward; that is, by working in a project, the project is yours--it gives you an opportunity to own it in some way. Second, participation gives you a chance to have something that isn't available to anyone else. A program in the early stages of field testing is available to an evaluation agency but not to a regular member of the consortium, and since it's available to the agency, it can serve a political purpose. It allows you to show it to superintendents, it allows the agency to work with the people who may, in the future, object to the project. Also, one can begin developing utilization activities given access to a program earlier.

One of the unique byproducts of formative evaluation and field testing in our agency has been that in a sense we are training formative evaluators at local education agencies because of their participation in our project. Many people who have worked with us in formative evaluation have later gone on to enter the evaluation field in graduate study or in regular job activities.

Formative evaluation is integral to our agency. Our activities and the distinctions between our beliefs and demands for formative data and that of other projects will probably be evident to you now.

Discussion of Formative Evaluation of Children's Educational Television Programs

Joanne Linowes
Massachusetts Educational Television

We have just heard about four types of evaluation efforts, both internal and external, for different kinds of children's educational television programs. Evaluations themselves may yield interesting data, but it is important to look at how these evaluations are utilized by persons or agencies making programming and production decisions. I shall address two areas of concern: 1) what role does evaluation play in program selection at the local level, and 2) what constitutes a successful interaction between evaluation and production.

First, programming decisions. Massachusetts Educational Television, as the state television agency, is responsible for the daytime broadcast of instructional television programs on the two public television stations in the Commonwealth. Its major objective is to provide high quality programming for instructional purposes which meets curriculum needs as determined by the State Department of Education and local educational agencies. As MET is primarily a distribution agency, as opposed to a production house, the schedule of broadcast programs is determined by a process of screening hundreds of available programs and selecting those which most closely meet stated educational needs. Programs are selected for their curriculum appropriateness, visual appeal, and ability to successfully transmit information to the intended audience. These programming decisions are based on the opinions of adult education and television specialists. No testing

is conducted. Rarely are students surveyed. The major assumption is that these programs have already been field tested and have been found to be educationally valid and visually appealing to the audience before they are released by the producers.

However, that is most often not the case. Why should producers devote a portion of their hard-to-come-by production monies to evaluation? Why should they have to put their productions on the firing line? Is it primarily because, as described in the Rebop case, evaluation was required in order to obtain funding? There may be little other incentive unless evaluation of an ETV production directly affects the programs' marketability.

How would this work? At MET, decisions for selection and then continuation of broadcasts are made as I just stated, first, by assessing curriculum needs and seeking appropriate programming from producers which fulfills these needs. Then, once the programming is broadcast, a summative evaluation is conducted to determine which subject areas, grade levels, and geographic locations are viewing which programs, and how teachers are using each series. Due to extremely limited funding available to local ITV agencies for on-going evaluation, and the great number of programs that need to be evaluated, this type of summative evaluation consists simply of teacher and student surveys. Teachers respond to these surveys based on their students' reactions to the programs; student learning from the program is not measured. Decisions as to whether or not to keep a series on the air are the result of this teacher and student input. If the

programs are not well received, they are dropped, regardless of their importance in the curriculum . . . and the producer is left with one less outlet for his/her productions. If, during production, the producer was to engage an evaluation team to conduct formative evaluation of the programs, he would have a more reliable product and the broadcast agency would make programming decisions based on hard core data rather than on the emotional appeal of the programs to teachers and students.

Since local broadcast agencies are unable to do extensive testing themselves, they should demand that programs be tested and revised to ensure they are obtaining the highest quality programming. Demand for evaluation must be made by the broadcasters; evaluators alone will have a difficult time convincing producers of the need for evaluation.

However, as we heard in the Rebop and Infinity Factory reports, just because formative evaluation has been conducted there is no guarantee that it will have a significant impact on production. There must also be successful interaction between the evaluation and production teams.

Each of the presenters referred to the three factors which affect the quality of the evaluation effort as it relates to production. These are: the production budget, the production time line, and the ability of the evaluation team to translate findings into specific production suggestions. Whether the evaluation strategies are designed to placate the funding source, it is the evaluator's responsibility to work with the production team at the outset of the production process to ensure

that their input will be incorporated into the programs. Put simply, the production budget must include adequate funds, not only for formative evaluation, but also to allow for re-shooting of segments which, when tested, fall below desired levels. The production time line must be developed so as to include program revisions which become necessary based on evaluation feedback. Those are the responsibilities of the producer. The responsibility of the evaluator is not only to report findings to the funding source and project advisors, but also to restate their findings so that they are specific to production persons. To say that a certain percentage of students don't watch a certain segment is not very helpful to a filmmaker. He needs to know much more detail and to work with the evaluation team to determine a more suitable approach. If the producer has allowed for adequate time and monies to incorporate the suggestions, the program will be improved. If not, evaluation has been of little value.

Evaluations conducted on A. I. T. programs are a good example of how the process can work. Nearly complete programs are released across the country for the purposes of testing. Certain states are asked to field test several programs. Massachusetts tested several programs in the Bread and Butterflies series. In one of the programs, "Decisions, Decisions," a segment was particularly didactic, and when tested, was found to have very few viewers paying attention. When this finding was substantiated in several other states, A. I. T.'s evaluators reported this to the production team. Since the programs were considered produced primarily for field test purposes, and since revisions had been budgeted

for, in terms of both time and money, the segment was re-shot in a visually more appealing way. Again the program was tested. The results indicated higher attention level and the new segment was incorporated into the program.

The Infinity Factory evaluation was less successful since most of the testing was done after nearly the entire series was produced. Thus, no matter how significant were the findings, data collected could have little or no effect on the individual programs tested, let alone on the remaining programs in the series.

In developing the original proposal for the new Native American series, the evaluation scheme was designed simultaneously with the production schedule. It allows for pre-production input and post-pilot segment testing as parts of the formative evaluation. Evaluation was seen as an integral part of the production process. Objectives for each program are to be guided by evaluators so that testing of the pilot can be carried out along formal guidelines. Reports to be prepared for the funding source will contain all of the formative evaluation results, but separate reports will be prepared for and discussions will be held with the production team. As designed in the proposal, it appears to be a workable system. However, to date, the production team is already doing content research and preparing to write script treatments, and the evaluators are barely on board. Will this project, too, fall into the pattern of strained relations between production and evaluators?

In conclusion, in order for evaluators to have a significant role in the development of children's educational television programs, findings must be incorporated into the productions. For this to happen, producers must plan for the inclusion of changes in their productions and evaluators must be able to effectively translate their findings into production terms. If the programs go through this production-testing-revision process, then the local broadcast agencies can select programs with the confidence that not only do the programs meet stated educational objectives, but also achieve particular learning objectives, and do both while successfully using the medium of television.

Discussion following symposium presentations

Question: Given the concern about insufficient funds, I'd like to ask why there isn't attempt made to approach commercial networks where there is alot more money to support evaluation of children's programs.

Valerie Crane: Joanne Linowes was approached by a local commercial television station in Boston to evaluate a preschool program. She asked me to assist her in that evaluation effort. The budget for the program was relatively low, the programs were all completed and yet we field tested a pilot program. There was no budget for this evaluation--MET performed the evaluation as a service oriented instructional television agency. However, the commitment to this kind of evaluation has in the past been relatively limited. That should change.

Question: I am not talking about approaching the networks. I am suggesting to approach companies who produce programs and then sell them to the networks. There I think the budget might allow for evaluation since the evaluator can provide the producer with data which will help them sell their product to the networks.

Saul Rockman: Most decisions about children's programs are made by adults and as long as they continue to be made by adults they'll be made on adult criteria. Knowledge of how children relate to a program and how much they learn from it is incidental to how many children are willing to watch and for that you have to understand the variety of concerns that go into commercial television and you're talking about scheduling concerns, you're talking about production quality, you're talking about action as opposed to education and those concerns may not be as appealing to many of us working in educational TV as they are to others. Now there are commercial test houses that do work on children's programming. There are a number of people in New York City who work as private consultants and who do a great deal of work on children's programming but much of it is market testing rather than content testing.

Frank Harvey: There has been some small testing of pilot programs funded by commercial people who are considering sponsoring a series of programs but once somebody buys the pilot, and you have the money up front, there seems to be no interest on the part of producers to ask for more money for ongoing evaluation. The producer says, "We know its going to be good. We have our instincts and we'll go ahead with them and we know everything we need to know about kids, how they think, what they like and what they don't like."

Question: I found the most interesting part of this discussion was the assessment of politics in these programs. Let's say we are here five years from now when we get together at another AERA symposium on evaluating TV programs in terms of what we've learned from this early history of TV programming. I'm just wondering what kind of documentation might be available now so that other evaluators won't have to start from scratch. I'm not only talking about instruments. Are there any kinds of documentation going on now?

Saul Rockman: We're working on a number of concerns that we have about understanding the whole impact of school television and its programming. The formative and summative evaluation reports for just about every project we know about are finding their way into ERIC. However, there is some concern from the other point of view, from the political point of view, from the decision making point of view. We have commissioned a study that was conducted last year and is yet to be formally reported on the concerns of the decision makers about the use of TV in the schools. What are the concerns that they have? What kind of data would they find useful in making their day to day decisions? That data I hope will be available soon.

We're also beginning to conduct some case studies of how television is actually being used in the schools. That is more than just "here is some data to show student achievement." What is the ongoing process of use? These are things we've never known.

On top of that and something that should have been done years ago is a utilization study being conducted by the Corporation for Public Broadcasting. We may be the only major educational field that doesn't know if its reaching anybody because we don't have any audience figures. This too will add some understanding of what TV is doing and how it is being utilized. These will come in time and in five years you'll have something.

Question: Aside from the direction that's being taken by AIT, I got the distinct impression that the evaluation attempts of these programs that are being created tend to look at these programs in isolation. For years we have been trying to teach teachers that programs don't stand alone but only work when they are previewed by the teachers, the students are prepared for them, and the teachers conduct some kind of follow-up. How can you justify the type of evaluation you are doing outside the naturalistic setting?

Valerie Crane: The ESAA funding is designed for home viewing (as stated in the RFP) and I think that the interest is much heavier in that area than in the area of in-school use. They want prime time programming for children and they are very concerned that this programming be prepared for viewing in the home. Therefore, the program must stand on its own. We do sometimes go into schools and evaluate these projects. I would question the validity of this procedure because the real concern is with home viewing without the teacher and school support.

Question: We just finished the evaluation on Vegetable Soup and we do have our program on both channels 13 and 4. I think we are one of the two programs that has two kinds of spots. The other thing that I found interesting today, interestingly missing, (and I realize at the end that it was not that kind of discussion group) and that is some knowledge of the instruments of the procedures used for collecting data. I think that that would be a very important thing to share with each other. We have our instruments and we certainly would like to share them with other evaluators who have done this kind of testing and we'd like to have other evaluators share this kind of thing with us. So that when we go to do it again we are a little more sure of what people are doing and can make revisions in that way also.

Bob Consalvo: A very significant need in the field is the sharing of instruments and research on the instruments themselves to see what effect they have or what kinds of data we can get with what kinds of instruments. Some of the participants are working on that problem right now and that was the intent of this symposium was to bring these evaluators together to start to talk about these issues. I'd suggest that's a very valid issue for a symposium.

Valerie Crane: Some of the people on this symposium have been invited to attend a NEERO meeting in May in Manchester, N.H. and we're encouraging anybody here involved in the evaluation of television to come. We are going to prepare recommendations for research and evaluation in children's television focusing on two areas. Basically we're interested in research which will have input into program development both in developing content and also in the role of evaluation. We hope to come up with a series of recommendations that we can submit to funding agencies to make people aware of the need for examining the many, many problems in evaluation. We want to progress beyond the stage where we just say, "I used this kind of a questionnaire, what did you use?" We want studies which will produce substantive kinds of data so we can become more effective evaluators.

Question: I'd like to give some encouragement to people on the panel and maybe some others in the room, to consider putting together now several symposia for next year because I think a number of issues have been identified right here in today's discussion. They could be addressed in several separate symposia. We talked about methodological issues of methods in measurement; we talked about a major issue of interfacing between evaluator and producer and I think that could have four or five experienced-based talks on how people have made the linkage with the producer. These are issues that certainly aren't written down with the exception of some of the Children's Television Workshop things which are suggestions how that might be effectively done. I'd like to give some encouragement to everyone to think about putting together a symposium right now.

Bob Consalvo: It may be a good idea for those who may be interested in forming some kind of network if we could get some names and addresses of people on a piece of paper up here at the end and we'd be glad to put it together and mail it to everyone and maybe we could begin to talk like that.

Saul Rockman: There's another organization involved in that, and that's the National Association of Educational Broadcasters. There is a research group within this association with its own newsletter and that often deals with these very issues.